Near-Complete Genome Sequences of Vesicular Stomatitis Virus Isolates from the 2020 Outbreak in Kansas

Tyler Doerksen,* Edward Bird,* Jamie Henningson,* Rachel Palinski*

*Kansas State Veterinary Diagnostic Laboratory, Kansas State University, Manhattan, Kansas, USA

ABSTRACT
Here, we report the near-complete genome sequences of vesicular stomatitis virus (VSV) serotype Indiana isolates from the 2020 U.S. outbreak. The sequences were obtained from swabs collected from Kansas horses in July and August. The four genome sequences help improve our understanding of VSV outbreak dynamics in the United States.

Vesicular stomatitis virus (VSV; genus Vesiculovirus, family Rhabdoviridae) is a negative-sense, single-stranded RNA virus that commonly causes vesicular disease outbreaks in cattle, sheep, goats, pigs, and horses in the Western Hemisphere. Vesicular stomatitis (VS), caused by VSV, manifests with vesicles around the mouth, tongue, ears, and coronary band (1, 2). Since the clinical presentation is indistinguishable from that of other high-consequence vesicular diseases, such as foot-and-mouth disease, rapid detection and diagnosis are essential. The two major VSV serotypes, Indiana (VSIV) and New Jersey (VSNJV), cause cyclical epidemics in the United States thought to originate from regions of Central America where VSV is endemic (3–5). The most recent U.S. VSV outbreak started in 2019 and affected 7 states (Colorado, Nebraska, New Mexico, Oklahoma, Texas, Utah, and Wyoming) that year, expanding unusually far east the subsequent year (2020), affecting 8 states in total (Arizona, Arkansas, Kansas, Missouri, Nebraska, New Mexico, Oklahoma, and Texas) (6).

Following the 2020 outbreak of VS in Kansas, lesion swabs were collected from horses with clinical signs of vesicular disease throughout southeastern Kansas and used in this study.

At Kansas State Veterinary Diagnostic Laboratory (KSVDL), RNA was extracted using the QiaAmp viral RNA minikit (Qiagen), followed by first-strand cDNA synthesis with random hexamers (Superscript III; Invitrogen) and second-strand synthesis with a non-directional polymerase (NEBNext Ultra nondirectional RNA second-strand synthesis module) (7). Primers were removed (HighPrep PCR cleanup system; MagBio), and libraries were prepared using the Nextera XT dual-indexing kit. Libraries were run on the Illumina MiSeq platform using a 300-cycle v2 cartridge. All sample preparation steps were performed following the manufacturer’s protocols.

Raw reads were trimmed, de novo assembled into contigs, and analyzed with BLASTn. The closest reference was identified as VSIV isolate IN0919COB (GenBank number MT437285), collected in 2019 from a bovine in Colorado (8). Trimmed reads were aligned to IN0919COB, and a consensus sequence was extracted. All bioinformatics steps were performed using default parameters in CLC Workbench v 20.0.3. The genomes were designated IN0720KSE, IN0720KSE2, IN0720KSE3, and IN0820KSE. Sequencing metrics are specified in Table 1.

We report the first VSV genome sequences from the 2020 U.S. outbreak season. The four genomes have a GC content of 41.8% and five open reading frames (N, NS, M, G, and L) encoding proteins of the expected sizes (423, 266, 230, 512, and 2,110 amino acids) (9, 10). The genomes are 99.90% to 99.99% identical to each other and 99.75%
to 99.81% identical to the IN0919COB reference, correlating to 1- to 11-nucleotide (nt) and 22- to 28-nt differences, respectively. Nucleotide differences between the samples are present in N (n = 2), G (2), L (6), and noncoding regions (1). Compared with the reference, nucleotide differences occur in N (2), NS (1), M (1), G (0), L (13), and noncoding regions (10).

Advances in deep sequencing technologies have allowed rapid sequencing of whole genomes concurrent with outbreaks. These advances present unique opportunities to analyze and predict outbreak dynamics in real time, thereby governing a subsequent response. The four near-complete genome sequences discussed in this study will fill key knowledge gaps associated with the molecular evolution of the most recent (2020) VSV outbreak.

Data availability. The consensus genome sequences were deposited in GenBank under accession numbers MW373776, MW373777, MW373778, and MW373779. The deep sequence reads were deposited in NCBI Sequence Read Archive (SRA) under project number PRJNA685396. This project references the first version of the sequences.

ACKNOWLEDGMENTS
We thank the Kansas Division of Animal Health in the Kansas Department of Agriculture and the Animal and Plant Health Inspection Veterinary Services in the United States Department of Agriculture for sample collection.

This project was funded internally by the Kansas State Veterinary Diagnostic Laboratory.

REFERENCES

### TABLE 1 Sequencing metrics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Data for strain:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IN0720KSE</td>
</tr>
<tr>
<td>No. of reads</td>
<td>418,744</td>
</tr>
<tr>
<td>Avg read length (bp)</td>
<td>136</td>
</tr>
<tr>
<td>No. of contigs</td>
<td>444</td>
</tr>
<tr>
<td>Avg coverage (%)</td>
<td>30.6</td>
</tr>
<tr>
<td>Genome length (bp)</td>
<td>11,051</td>
</tr>
<tr>
<td>% similarity to reference</td>
<td>99.81</td>
</tr>
</tbody>
</table>